E. Probe Mapping

Table S5. Summary of Probe Mapping per Gene Expression Platform

Platform	Length of mapped probe sequence[a] (nt)	Number of probe sequences analyzed[b]	Number of probes that met mapping criteria[c] (percent of all probes, %)		Number of RefSeq NM Accessions mapped to probes[e]	Number of Entrez genes ID's mapped to probes via		Number of genes, not in Entrez, mapped to probes via AceView[d,f,g]
			RefSeq	AceView[d]		RefSeq[e]	AceView[d,f]	
ABI	60	32,878	18,547 (56.4)	25,566 (77.8)	21,963	16,763	18,676	3,267
AFX	25	54,675	24,694 (45.2)	44,693 (81.7)	21,318	15,965	18,911	10,129
AG1	60	41,000	22,677 (55.3)	32,024 (78.1)	21,890	16,493	18,051	4,055
GEH	30	53,423	16,881 (31.6)	43,540 (81.5)	20,230	15,429	16,984	18,408
ILM	50	47,282	20,140 (42.6)	31,229 (66.0)	22,161	16,990	18,797	8,666
NCI	39-70	35,235	21,555 (61.2)	29,396 (83.4)	20,987	15,899	17,641	1,411
EPP	161-513	294	285 (98.6)	285 (98.6)	315	285	290	0
QGN	183-2,671	245	234 (95.5)	234 (95.5)	253	233	237	0
GEX[h]	N/A	205	N/A	N/A	203	203	203	N/A
TAQ[h]	N/A	1,004	N/A	N/A	997	997	997	N/A
Union of six platforms[i]		264,493	125,216 (47.3)	206,448 (78.1)	23,971	18,114	21,662	32,025
Intersection of six platforms[i]					15,615	**12,091**	13,327	9
Intersection of six platforms[i] *and TAQ*[h]						**906**		

[a]For the AFX platform, the length of each individual probe is given. For the QGN platform, the length of the intended target is given.

[b]The number of probes for which mapping was attempted may slightly differ from the number of probes arrayed (**Table 1**) because of the removal of control probes and replicate spots. For the AFX platform, the number of probe sets is given.

[c]Probes were mapped as described in the Methods section. An exact sequence match was required and probes that match more than one gene were excluded. For the AFX platform, there are generally 11 probes per probe set, and each probe was mapped individually. An exact match of 80% of the probes in a probe set was required for the probe set to qualify as a perfect match. All the mapping data supporting this table are available from supplementary materials online and the MAQC web site (http://edkb.fda.gov/MAQC/).

[d]AceView is a transcriptome database that combines RefSeq, GenBank and dbEST entries [Thierry-Mieg, D & Thierry-Mieg, J, Genome Biology **7** (Suppl 1):S12**,** 2006]. For the details on the AceView mapping, please refer to the supplementary materials online at ftp://ftp.ncbi.nlm.nih.gov/repository/acedb/MAQC/MaqcMapping2AceViewTranscripts.zip.

[e]The numbers in these columns illustrate the source of the common set of 12,091 genes represented on the six high-density microarray platforms which have an overlap of 906 genes with the TAQ platform. The data do not fully reflect the coverage of each platform because the degree to which RefSeq and non-RefSeq sequences are emphasized during probe design and selection differs among the platforms.

[f]The number of Entrez genes specifically assayed, through any of their alternative transcript variants, is given in these columns. Probes with a few gaps or mismatches were permitted, but at the same time, probes with even a minor risk of cross-hybridization to another gene (with up to 30% mismatches) were ignored.

[g]Genes, not yet in Entrez, are supported by cDNAs in GenBank, and are described in AceView. The sum of genes in Entrez (via AceView) and genes not in Entrez that mapped to probes (via AceView) is the total number of genes in the AceView database that are matched by each platform under the mapping criteria chosen for this study.

[h]For the two PCR-based platforms (GEX and TAQ), no exact sequence mapping was conducted. Consequently, assay annotation information provided by the manufacturers was used to determine cross-platform mapping.

[i]The union and intersection numbers are based on the six high-density microarray platforms (ABI, AFX, AG1, GEH, ILM, and NCI).